



Impact of Induced Fit on Ligand Scoring and a Strategy of Identifying a Minimal Set of Flexible Residues

D. Kokh, W. Wenzel

published in

*From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,*
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 257-260, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Impact of Induced Fit on Ligand Scoring and a Strategy of Identifying a Minimal Set of Flexible Residues

Daria Kokh¹ and Wolfgang Wenzel²

¹ FB.C - Mathematik und Naturwissenschaften, Bergische Universität Wuppertal,
42097 Wuppertal, Germany
E-mail: kokh@uni-wuppertal.de

² Institute für Nanotechnologie, Forschungszentrum Karlsruhe GmbH,
Postfach 3640, 76021 Karlsruhe, Germany
E-mail: wenzel@int.fzk.de

Although conformational changes in receptor upon ligand binding are a very common phenomenon, incorporating protein flexibility in a docking procedure encounters significant computational problems. A possible solution is inclusion side-chain flexibility for only limited number of residues in the binding pocket, which can improve notably docking accuracy without considerable increase of computational costs. However, investigation of this approach is often limited to specifically chosen receptors and mostly focused on the impact of receptor flexibility on docking accuracy, whereas ligand scoring, the real weakness of the present-day docking methodology, is treated only peripherically. In the present study we investigate enrichment rates of rigid-, soft-, and flexible- ("induced-fit") -receptor models using 12 diverse proteins with receptor-specific ligand libraries containing up to 13000 molecules, comprising known ligands and decoys with similar physical properties but distinct topology. We also present and test a straightforward protocol for the choice of the flexible residues, which is based on the ability of the receptor structure to accommodate the set of known ligands. This strategy is an unbiased approach to identify the most important residues likely to be relevant for induced fit effects, which allowed us to improve EF₁ values by ~35% on average with respect to rigid-docking.

1 Method

FlexScreen¹ is an all-atom docking approach based on the stochastic tunneling method² for the energy minimization and a simple, first principle based atomistic scoring function that contains a sum of the Van-der-Waals, electrostatic Coulomb, and angular dependent hydrogen bond. The Van-der-Waals parameters have been taken from OPLSAA³, the partial charges of the receptors have been computed with MOE⁴, the hydrogen bond parameters have been taken from AutoDock⁵. The method enables continuous rotation up to 15 side chain bonds of the receptor in the energy optimization procedure.

Scoring performance of the Flex-Screen approach has been benchmarked by using 12 target proteins of the DUD database⁶ with relatively small binding cavities that are completely buried from solvent. For each target the database includes a set of annotated ligands (up to 350) and a set of decoys containing about 36 molecules for each ligand that resemble the particular ligand in physical properties, but differ topologically, so that they unlikely to be binders. The following receptor have been analyzed: Androgen receptor, Cyclooxygenase 1, Cyclooxygenase 2, Estrogen receptor agonist, Glycogen phosphorylase beta, Glutacorticoid receptor, Mineralcorticoid receptor, Purine nucleoside phosphorylase, Progesterone receptor, Retinoic X receptor alpha, S-adenosyl-homocysteine hydrolase, and Thymidine kinase.

Docking and screening performance has been evaluated by computing the enrichment of annotated ligands among the top-scoring molecules of the receptor-specific database

$$EF_{\alpha} = \frac{(\text{concentration of known ligands found in top - ranking subset})}{(\text{concentration of known ligands in database})} \quad (1)$$

where EF_{100} equals the fraction of annotated ligands that bind to the receptor in the docking calculations (binding energy less than zero) and, therefore, shows the efficiency of ligand docking, whereas EF_1 indicates screening efficiency with the maximum possible value of 37.

In order to treat receptor flexibility for all systems at an unbiased level, some uniform scheme for choosing the flexible residues has to be implemented. In the present study potentially flexible residues have been identified by requiring that majority of the known ligands must have a negative binding energy to the receptor. We have, therefore, analyzed a list of ligands that are unbound in rigid-receptor docking for a specific target and found a set of residues that most often cause energy clashes with these ligands (vdW energy above 20kJ/Mol). These residues have been ranked according to the number of clashes and form a list of flexible residues. Finally, several top-ranking residues from this list have been treated as flexible.

2 Results

We have found significant limitations of rigid receptor models that for some targets fail to bind even 50% of the known ligands to the apo-structure of the protein. The enrichment rate does not correlate with docking performance and is good only for 4 receptors :ER-agonist, COX2, MR, PNP (see Fig.1). For the other receptors enrichment rates remain poor, in high correlation with a previous study of the same database⁶.

We have investigated a soft- and flexible-receptor approximations, by shifting or optimizing (soft- and flexible-docking, respectively) of side-chains from the list of flexible residues to adapt binding pocket for annotated ligands.

As can be expected, both models are effective in finding binding poses for ligands that do not dock in the rigid-receptor calculations, whereby the fraction of annotated ligands that bind to a receptor (EF_{100}) increases monotonically with the number of shifted/flexible residues as illustrated in Fig.2. The variation of enrichment rate, however, is not monotonic and reaches its maximum at about 3-8 flexible/shifted residues for major receptors. Since the energy correction accounting for receptor reconstruction is omitted in soft-receptor model, it is not surprising that this method is not so successful with regards to the enrichment performance: EF_1 do not even reach the values obtained in rigid docking for most receptors (Fig.1, left panel). Unlike soft-receptor, flexible-receptor model increases enrichment rates in comparison with the rigid receptor model for 11 from 12 targets (Fig.1, right panel). The scoring performance of flexible-receptor docking is good ($EF_1 > 20$) for 8 of 12 targets (in comparison to 4 in the case of rigid-docking) and medium ($10 < EF_1 < 20$) in the remaining 4 cases. In contrast to rigid-receptor docking, where for 4 targets screening results are unsatisfactory, we now find $EF_1 > 10$ for all targets.

These results show that accommodation of ligand-induced protein reconstruction by rotating of receptor side chains that are most often involved in steric clashes between a protein and known ligands can notably improve performance of visual screening. The

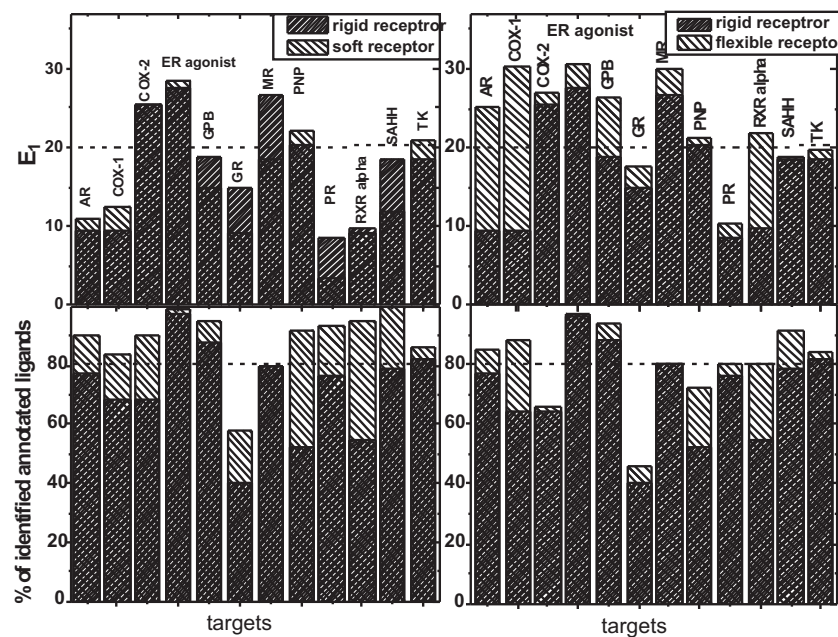


Figure 1. Comparison of enrichment factors of rigid- and soft- (flexible)-receptor models for 12 receptors. The numbers of flexible residues are: 8 (AR), 9 (COX1), 3 (COX2), 4 (ER-agonist), 6 (GPB), 6 (GR), 6 (MR), 3 (PNP), 7 (PR), 11 (RXR-alpha), 2 (SAHH), and 3 (TK). In soft-docking selected residues are shifted by 0.25nm.

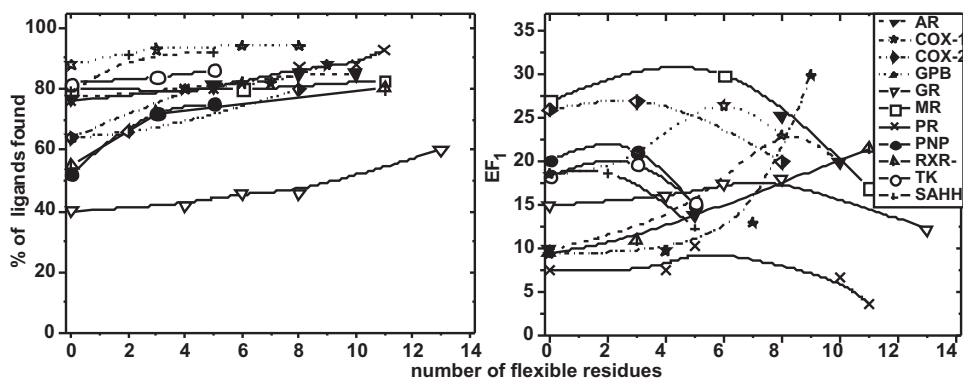


Figure 2. Variation of enrichment factors with the number of flexible residues employed in flexible docking.

present algorithm does not require empirical assumptions of possible soft spots of the receptor binding pocket. Instead, accuracy of the method depends mainly on the number and variety of ligands known to be bound to the specific target. If only a limited number of known ligands is available, compounds with similar physical properties may be used to explore the active site and create a list of flexible residues.

References

1. H. Merlitz and W. Wenzel, *Comparison of of stochastic optimization methods for receptor-ligand docking*, J. Chem. Phys. Lett. **362**, 271-277, 2002.
2. W. Wenzel and K. Hamacher, *Stochastic tunneling approach for global optimization of complex potential energy landscapes*, Phys. Rev. Lett. **82**, 3003-3007, 1999.
3. W. L. Jorgensen and N. A. McDonald, *Development of an all-atomic force field for heterocycles properties of liquid pyridine and diazenes*, J. Mol. Struct. **424**, 145, 1997.
4. *Molecular Operating Environment (MOE) version 2003.02*; Chemical Computing Group Inc.:Monreal, Canada, 2003.
5. G. M. Morris, D. S. Goodsell, R. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding energy function*, J. Comput. Chem. **19**, 1639, 1998.
6. N. Huang, B. K. Shoichet, and J. J. Irwin *Benchmark sets for molecular docking*, J. Med. Chem. **49**, 6789-6801, 2006.
<http://blaster.docking.org/dud/>.